The effect of grade retention on school outcomes: an application of doubly robust estimation method

Uysal, Selver Derya

Center for Quantitative Methods and Survey Research

CMS Discussion Paper 2010-01

The Effect of Grade Retention on School Outcomes: An Application of Doubly Robust Estimation Method

Selver Derya Uysal*

*University Konstanz, Department of Economics

http://www.uni-konstanz.de/cms

The Effect of Grade Retention on School Outcomes:
An Application of Doubly Robust Estimation Method

Selver Derya Uysal*

Department of Economics

University of Konstanz

January 31, 2010
PRELIMINARY VERSION

Abstract

In this study, I estimate the average causal treatment effect of grade retention on several educational outcome variables, such as completion of upper secondary school, graduation grades in math and German, as well as average final grade using a data set from Germany. The analysis relies on Conditional Independence Assumption. I use doubly robust method, regression adjustment and inverse propensity score weighting. The results of the empirical study show that grade retention does not improve the students' educational achievement.

JEL classification: $C21, I2$

Keywords: Econometric evaluation, doubly robust estimation, conditional independence assumption, average treatment effect, propensity score, grade retention

*Department of Economics, Box D124, University of Konstanz, 78457 Konstanz, Germany. Phone +49-7531-88-2556, Fax -4450, email: selver.uysal@uni-konstanz.de. The author gratefully acknowledges financial support of the German Science Foundation (DFG) through the Research Group Heterogeneous Labor

1 Introduction

Grade retention is an intervention tool in education. It refers to the practice of requiring a student to repeat the same grade which s/he has already completed because of her/his poor performance. In Jackson (1975), the aim of grade retention is explained as an attempt at remedying inadequate academic progress and contributing to the development of students not ready for the next grade. The underlying idea is that students who do not successfully complete a grade level will not be able to digest the next higher grade's material. These students are therefore, for their own interest, required to repeat the grade. The most important question, however, is whether grade retention really helps students to improve their grades or whether it harms the students' school success. This paper aims to address this question and estimates the causal effect of this school intervention on several school outcomes.

The effects of grade retention have been a discussion topic for more than four decades. Most studies concentrate on the effects of grade retention on performance in later grades, on the likelihood of drooping out of high school, and on labor market outcomes for late adolescence (see Guevremont, Roos, and Brownell (2007), McCoy and Reynolds (1999a), Jimerson (1999), Jimerson (2001), and Eide and Showalter (2001) among others.). The results are somewhat controversial: although the vast majority of empirical work done with the data from the US and Canada points out the negative effects of grade retention, there are also a number of papers indicating gains.

Since being held in a grade is not a random assignment, simple mean comparisons of outcome variables do not reveal the true causal effect of grade retention. We could realize true causal effects over a whole population by using mean comparisons, if we could randomly hold schoolchildren in the same grade for a second year. Since such an experiment on schoolchildren is impossible and unethical, we should rely on the econometric methods which enable identification of the true causal effects in terms of potential outcomes. In the case of binary treatment, there are two potential outcomes for treated and nontreated cases: one observed depending on the realized treatment status, and the other one unobserved (i.e. counterfactual). Identification is achieved under some assumptions in potential outcome framework. The crucial assumption I am using in this paper is Conditional Independence Assumption[1]. It means given a set of observable characteristics which are not affected by the treatment, potential outcomes are independent of treatment assignment. There are several methods proposed for estimating treatment effects under the assumption of conditional independence (see Imbens (2004) for a review). The main methods can be categorized into regression, propensity score weighting and matching methods. Here, I estimate the effect of grade retention on different outcomes using regression, propensity score weighting and a combination of regression and propensity score weighting methods. The advantage of the combination over the single methods is that the mixed method provides double protection against misspecification. That is, the estimator is still consistent, even if either the propensity score or the mean function is wrongly specified but not both (for further discussion of double robustness see Robins, Rotnitzky, and Zhao (1995), Robins and Ritov (1997), Hirano and Imbens (2001), Wooldridge (2007), and Bang and Robins (2005)).

In this paper I use a German dataset "Gymnasiastenstudie" (Central Archive for Empirical Social Research (2007)) in order to estimate the causal effect of grade retention on different school outcomes. This work distinguishes from the existing literature in many ways. First of all, to my knowledge, there is no empirical study published which analyzes the effects of grade retention using a Germany dataset. The dataset I use here is restricted to students attending upper secondary school (Gymnasium) in North Rhine-Westphalia. However, it is still representative for Germany, since one fourth of the German population resides in North Rhine-Westphalia and it is the biggest federal state in terms of population among the 16 federal states in Germany. Furthermore, one forth of the students in Germany is attending school in North Rhine-Westphalia. Besides that the upper secondary schools (Gymnasien) in Germany serve almost for one half of the total students after primary education (Grundschule)[2]. This paper is also one of the very few papers which rely on econometric evaluation methods in order to analyze the effect of grade retention on school outcomes. Another contribution of this paper is that it uses one of the least applied econometric evaluation methods, namely Doubly Robust Method[3].

[1]This assumption is called *Ignorability of Treatment* (given observed covariates X) by Rosenbaum and Rubin (1983) and *Unconfoundedness* by Imbens (2004).

[2]The exact numbers can be found on the website of Federal Statistical Office: http://www.destatis.de/.

[3]To my knowledge, there are only two applications of this method: Bang and Robins (2005) and Hirano and Imbens (2001)

The organization of the paper is as follows: Section 2 gives a short review of the existing literature. Section 3 briefly explains identifying assumptions and the econometric methods applied. Section 4 focuses on the sample and elaborates on the empirical results. Finally, Section 5 summarizes the main results and concludes the paper.

2 Literature Review

Grade retention has been an important topic in the last four decades especially for educational researchers. This research is concentrated on characteristics of the students who repeat a grade and the effect of grade retention on different outcomes, such as academic achievement, socioemotional outcomes, behavioral outcomes and employment outcomes. The studies from educational research can be characterized as more explorative data analysis rather than causal analysis. Nevertheless, economists recently also show some interest on grade retention an its effects taking into account possible causality issues (see for example Greene and Winters (2009),Corman (2003), Eide and Showalter (2001),Jacob and Lefgren (2002), (2007) among others). Xia and Kirby (2009) give a very comprehensive overview of research done on grade retention.

The large body of literature dealing with the characteristics of retained students agree on many points. Most of the research show that boys are more likely to be retained than girls (for example Byrd and Weitzman (1994), Dauber, Alexander, and Entwisle (1993), El-Hassan (1998), Fine and Davis (2003), Guevremont, Roos, and Brownell (2007), Hong and Yu (2007), McArthur and Bianchi (1993), Frederick and Hauser (2008), Jimerson (1999)). Among others most of the above references also point out that the retained children come from families with lower socioeconomic characteristics such as low household income, lower educational attainment and lower occupational position. Parents of the retained students show on average less interest in their child's school education. These studies indicate that the retained students have lower cognitive skills (Blair (2001), Liddell and Rae (2001)), and lower noncognitive skills, such as self conception, self confidence or social competence (Ferguson, Jimerson, and Dalton (2001), Jimerson et al. (1997), Robles-Pina, Defrance, and Cox (2008)), compared to nonretained students.

The relation between grade retention and high school drop out has been investigated in education research a lot. Several studies show that retained students are more likely to drop out from high school (Allensworth (2005), Goldschmidt and Wang (1999), Guevremont, Roos, and Brownell (2007), Jimerson (1999), Jimerson, Anderson, and Whipple. (2002), Jimerson, Ferguson, Whipple, Anderson, and Dalton (2002), Roderick (1994)). Jacob and Lefgren (2007) use standard parametric regression discontinuity design to investigate the relation between grade retention and high school completion and they show that grade retention has a significant negative effect on high school completion for older students but the effect is insignificant for younger ones. Eide and Showalter (2001), however, show that the IV estimate of the effect of grade retention on high school drop-out is insignificant.

The results of studies on the effect of grade retention on academic achievement are somehow more controversial. Balitewicz (1998), Beebe-Frankenberger, Bocian, MacMillan, and Gresham (2004), Frymier (1997), Guevremont, Roos, and Brownell (2007), Hong and Yu (2007), Jimerson (1999), Jimerson (2001), McCoy and Reynolds (1999b) show the negative effects of grade retention on academic achievement using different methods. On the other hand, Greene and Winters (2004), (2007), (2009) show some positive effects of grade retention on academic achievement.

3 Econometric Method

Consider N units which are drawn from a large population. For each individual i in the sample, where $i = 1, ..., N$, we observe the triple (Y_i, D_i, X_i). D_i shows the binary treatment status for individual i:

$$D_i = \begin{cases} 1, & \text{if the } i^{th} \text{ individual is treated} \\ 0, & \text{otherwise} \end{cases}$$

We observe also a vector of characteristics (covariates) for the i^{th} individual denoted by X_i. For each individual there are two potential outcomes (Y_{i0}, Y_{i1}). Y_{id} denotes the outcome for each individual i, for which $D_i = d$ where $d \in \{0, 1\}$. For each individual only one of the potential outcomes is observed depending on the treatment status. The observed outcome, denoted by Y_i in the triple, can be written in terms of treatment indicator (D_i) and the potential outcomes:

$$Y_i = D_i Y_{i1} + (1 - D_i) Y_{i0}$$

Our primary interest lies in estimating the average causal effect of the repeating a grade. This effect is called the average treatment effect (ATE). It gives the mean effect of the treatment:

$$\tau = E[Y_{i1} - Y_{i0}] = E[Y_{i1}] - E[Y_{i0}]$$

Since only one of the potential outcomes is observed, ATE cannot be identified without further assumptions. For the empirical study I assume that the following assumptions hold:

Assumption 3.1 *Conditional Independence Assumption (CIA)*
$Y_{i0}, Y_{i1} \perp D_i | X_i$, *where* $\perp$ *stands for independence.*

It implies that after controlling for the effect of covariates, treatment and outcomes are independent.

Assumption 3.2 *Common Support*
$0 < Pr(D_i = 1|X_i) < 1$

Assumption 3.2 means that for all x there is a positive probability of either participating ($D_i = 1$) or not participating ($D_i = 0$). In other words for each value of covariates there are both treated and untreated cases. Thus, there is an overlap between the treated and untreated subsamples. If the assumption fails, then we could have individuals with x vectors who are all treated and those with a different x vector who are all untreated.

Rosenbaum and Rubin (1983) show that under CIA identification can be achieved by conditioning on a function of X_i, a balancing score[4], instead of a high dimensional X_i itself. The most commonly used balancing score in the evaluation literature is the propensity score, the conditional probability of assignment to the treatment given the covariates:

$$p(x) = Pr[D_i = 1|X_i = x] = E[D_i|X_i = x] \quad (3.1)$$

[4]A balancing score is a function of observed covariates X_i such that the conditional distribution of X_i given balancing score is the same for treated and control units (see Rosenbaum and Rubin (1983)).

Lemma 3.1 *Unconfoundedness Given the Propensity Score*
Given the CIA and Common Support assumptions, outcomes Y_{i0} and Y_{i1} are independent of treatment given the propensity score.
$Y_{i0}, Y_{i1} \perp D_i | p(X_i)$

Under these assumptions several methods can be used to estimate the average treatment effect. This paper uses three different methods: regression method, inverse propensity score weighting method and Doubly Robust Method which is the combination of the first two methods.

3.1 Regression

Under the CIA one can estimate the unconditional means $E[Y_{id}] = \mu_d$ based on the parametric estimation of conditional means $E[Y_{id}|X_i = x]$ for $d \in 0, 1$. Since the arguments are symmetric, I concentrate on $E[Y_{i1}|X_i = x]$. Assume that the conditional mean function is correctly specified, $E[Y_{i1}|X_i = x] = m_1(x, \beta_1)$, where $m_1(x, \beta_1)$ is a function depending on a covariate vector and a $k-$dimensional true parameter vector β_1. Given a consistent estimator $\hat{\beta}_1$, a consistent estimator of the unconditional mean, μ_1, is:

$$\hat{\mu}_1 = \frac{1}{N} \sum_i m_1(X_i, \hat{\beta}_1) \tag{3.2}$$

since $\mu_1 = E[m_1(x, \beta_1)]$ by iterated expectations.

Thus, one can estimate the average treatment effect based on two parametric regressions as follows:

$$\hat{\tau}_{reg} = \frac{1}{N} \sum_i [m_1(X_i, \hat{\beta}_1) - m_0(X_i, \hat{\beta}_0)] \tag{3.3}$$

From Wooldridge (2002) and Wooldridge (2009), the asymptotic variance can be written as follows:

$$
\begin{aligned}
AV\sqrt{N}(\hat{\tau}_{reg}) = & E[(m_1(X,\beta_1) - m_0(X,\beta_0) - \tau_{reg})^2] \\
& + E[\frac{\partial m_1(X,\beta_1)}{\partial \beta_1'}]V_1 E[\frac{\partial m_1(X,\beta_1)}{\partial \beta_1'}]' \\
& + E[\frac{\partial m_0(X,\beta_0)}{\partial \beta_0'}]V_0 E[\frac{\partial m_0(X,\beta_0)}{\partial \beta_0'}]'
\end{aligned} \tag{3.4}
$$

where V_1 and V_0 are the variances of β_1 and β_0. The variance can be estimated by replacing the expectations with the sample means and true parameters with their estimates.

3.2 Weighting by Propensity Score

Using Lemma 3.1, the mean outcomes for the treatment and control groups can be identified by weighting the observations with the inverse of the propensity score:

$$E[Y_{i1}] = E[DY/p(X)]$$

$$E[Y_{i0}] = E[(1-D)Y/(1-p(X))]$$

Hence, we can write the ATE as follows:

$$\tau = E[\frac{DY}{p(X)} - \frac{(1-D)Y}{(1-p(X))}]$$

The estimator of ATE can be written as a sample counterpart of the population expectation. Usually this estimator is referred as the propensity score weighting estimator[5]:

$$
\begin{aligned}
\hat{\tau}_{ps} &= \frac{1}{n}\sum_{i=1}[D_iY_i/p(X_i;\hat{\alpha}) - (1-D_i)Y_i/(1-p(X_i;\hat{\alpha}))] && (3.5) \\
&= \frac{1}{n}\sum_{i=1}\frac{(D_i - p(X_i;\hat{\alpha}))Y_i}{p(X_i;\hat{\alpha})(1-p(X_i;\hat{\alpha}))} \equiv \frac{1}{n}\sum_{i=1}\hat{g}_i && (3.6)
\end{aligned}
$$

Since usually the true propensity score $p(X)$ is not observable, one can use an estimated propensity score $p(X;\hat{\alpha})$, where $\hat{\alpha}$ is the maximum likelihood estimator (MLE) (e.g., probit or logit) of the parameter vector of the propensity score speci-

[5]This estimator is identical to an estimator from Horvitz and Thompson (1952) for handling nonrandom sampling.

fication. $\hat{\tau}_{ps}$ is inconsistent, however, if the propensity score is misspecified (see for further discussion Horvitz and Thompson (1952), Rosenbaum (1987), and Bang and Robins (2005))[6].

Following Wooldridge (2007), Wooldridge (2009) shows that the asymptotic variance of τ_{ps} is:

$$AV\sqrt{N}(\hat{\tau}_{ps} - \tau) = E[e_i e_i'] \tag{3.7}$$

where $e_i \equiv g_i - E[g_i s_i']E[s_i s_i']^{-1} s_i$, s_i is the score function of the MLE model of the propensity score.

3.3 Doubly Robust Method

Both of the above mentioned estimation methods, regression and propensity score weighting, can be easily implemented. There are no computational difficulties, or curse of dimensionality problems as in nonparametric methods. As mentioned above, consistency of the estimates hinges upon the true specification of the mean or the propensity score, depending on which estimation method is used. Wooldridge (2007) and Hirano and Imbens (2001) show, however, that combining weighting and regression methods gives a doubly robust estimate of the unconditional mean, providing double protection against misspecification. As long as one of the functional form specifications, either that for the conditional mean or the propensity score, is correctly specified, the resulting estimator for the unconditional mean will be consistent provided that $E[Y_d] = E[m_d(x, \beta_d^*)]$ where β_d^* is the probability limit of an estimator from the conditional mean function (Wooldridge (2007)). This property holds for linear exponential family with a canonical link function (see for details Wooldridge (2007), Scharfstein, Rotnitzky, and Robins (1999)). The three regression models I use for this application, namely linear, logit and poisson regression, belong to this family.

The main idea is weighting the objective function of the regression by the inverse of the propensity score. Depending on the choice of the regression method, the coefficient estimates of the mean function parameters come from weighted least square or weighted MLE method. The score function of the chosen parametric model is weighted by $1/p(X_i; \hat{\alpha})$ and by $1/(1 - p(X_i; \hat{\alpha}))$ for treated and untreated subpopu-

[6]Hirano, Imbens, and Ridder (2003) examine the estimator in equation 3.5 where $p(X_i; \hat{\alpha})$ is replaced by nonparametric estimates.

lation respectively.

Depending on the nature of outcome variable the proper mean function is one of the following:

- For a continuous outcome variable:

$$m_d(X_i, \beta_{dw}) = X_i'\beta_{dw} \tag{3.8}$$

- For a binary outcome variable:

$$m_d(X_i, \beta_{dw}) = \Lambda(X_i'\beta_{dw}) = \frac{\exp(X_i'\beta_{dw})}{1 + \exp(X_i'\beta_{dw})} \tag{3.9}$$

- For a count outcome variable:

$$m_d(X_i, \beta_{dw}) = \exp(X_i'\beta_{dw}) \tag{3.10}$$

The estimated coefficient $\hat{\beta}_{dw}$ from weighted regression method solves the weighted score function

$$\frac{1}{N}\sum_i w_i(Y_i - m_d(X_i, \hat{\beta}_{dw}))X_i = 0 \tag{3.11}$$

where

$$w_i = \begin{cases} 1/p(X_i;\hat{\alpha}), & \text{if} \quad D_i = 1 \\ 1/(1 - p(X_i;\hat{\alpha})), & \text{if} \quad D_i = 0 \end{cases}$$

Thus, one can estimate the average treatment effect based on two weighted regression coefficients as in regression methods:

$$\hat{\tau}_{dr} = \frac{1}{N}\sum_i [m_1(X_i, \hat{\beta}_{1w}) - m_0(X_i, \hat{\beta}_{0w})] \tag{3.12}$$

The asymptotic variance of $\hat{\tau}_{dr}$ is same as Equation 3.4 with different V_0 and V_1[7]. When estimating V_0 and V_1, one has to take into account that the weights are estimated in a first step. Wooldridge (2007) derives the asymptotic variance of $\hat{\beta}_{dw}$

[7]For the linear case the asymptotic variance of $\hat{\tau}_{dr}$ is equivalent to the variance derived by Hirano and Imbens (2001) for linear mean function.

as follows.

$$AV\sqrt{N}(\hat{\beta}_{dw}) = A_0^{-1}D_0A_0^{-1}$$

where $A_0 \equiv E[H(X, \beta_{dw})]$ and $D_0 \equiv E[k_i k_i']$. $k_i = k(X_i, \beta_{dw}) = w_i(Y_i - m(X_i, \hat{\beta}_{dw}))X_i$ is the weighted score function and $H(X, \beta_{dw})$ is the Hessian. Wooldridge (2007) proposes also the following consistent estimators for A_0 and D_0 :

$$\hat{A} = \frac{1}{N}\sum_i w_i H(X_i, \hat{\beta}_{dw})$$

$$\hat{D} = \frac{1}{N}\sum_i k(X_i, \hat{\beta}_{dw})k(X_i, \hat{\beta}_{dw})'$$

4 Data and Empirical Results

In the following, the causal effect of grade retention on several school outcomes is investigated for the German school system. The data set consists of information on family background and school related topics for about 3000 10^{th} grade students attending upper secondary school in North Rhine-Westphalia in the year 1970[8]. The students were sampled from 121 classes at 68 upper secondary schools. The data contains information from student, parent and teacher questionnaires. About ten years later, the students' grades were collected from the schools.

The empirical study on the causal effect of grade retention is distinguished from earlier studies by its investigation of the effect in a potential outcome framework and its application of the above explained methods for estimating the ATE of grade retention on school performance. Treatment is defined as repeating a class at least once after 10^{th} grade. The effects of grade retention on different outcome variables are investigated. The first one is the probability of graduating from upper secondary school (having "Abitur" or not). The other three outcome variables are only measured for those who have graduated from upper secondary school. One is the average final grade in upper secondary school. In addition, the effect on math and German "Abitur" grades is also considered. The aim of the empirical part is twofold: (i) estimate the causal effect of grade retention on the school performance, and (ii) investigate the differences of the causal effect for girls and boys. Outcomes are assumed to be independent of treatment status conditional on the covariates. All variables used in the study are listed in Table A1.

[8]The original data set consists of two more follow-ups in years 1984 and 1998.

The variables are chosen in accordance with earlier findings concerning characteristics associated with being retained as well as with being successful in school. It is important to include variables related to both treatment status and potential outcomes so that the CIA holds approximately. A female dummy is included because most studies show that males are more likely to be retained than females. A measure of intelligence, IQ, is also included to control for the cognitive skills of the students. The variable IQ in our study is the sum of correctly solved questions of a standard psychometric Intelligence Structure Test (IST), which was administered in the class-room in the 10^th grade. Since noncognitive skills also appear to play an important role in school performance, as shown in earlier studies, variables which measure the attribution of success to diligence (DILIG) and ability (ABIL) are included as conditioning covariates. The variable WISH is added as a control for the child's motivation. I also control for the age of the student. Former studies also claim that the characteristics of parents, such as economic well being, education and parental involvement with their child's school performance, are also likely to affect the probability of being retained. EDU_MOT, EDU_FAT, AGEMOT, HHINC, INTERSCHOOL are variables which control for family background and parents involvement. I can also identify whether the child has experienced any grade retention before 10^{th} grade (PR_RET).

The variables which are used in this study are chosen from three different sources. The outcome variables are taken from the administrative school data and the control variables are taken from parents and students questionnaires. Merging these three different data sets decreases the sample size already by about 500 observations. Some questions are asked to both students and parents. Thus, I combine the information sets to keep the decrease in the sample size moderate.

I created different samples. With the first sample I analyze the causal effect of grade retention on high school graduation (ABI) (see Table A2 for descriptive statistics). Thereafter, I restrict our sample for those who graduated from upper secondary school in order to estimate the causal effect of grade retention on average final grade (GPA) and final grades in math (MAT) and German (GER) (see Table A3 for descriptive statistics). Next I restrict the sample to the students who did not experience any grade retention before the 10^{th} grade (see Table A4) in order to see the effect of late grade retention on those students. For this sample I also look at the upper secondary school graduates and the effect of grade retention on graduation grades (see Table A5). For all four samples the analysis is done for the entire sample and for the subsamples by gender. The propensity score, the probability of being retained after 10^{th} grade is estimated by a logit regression for all subsamples. The regression results can be found in Table A6 and A7. Table A6 gives the logit estimation results for the sample before restricting by previous retention status and A7 gives the results only for students who did not experience retention before 10^{th} grade. From the logistic regression results, we can conclude that females are less likely to be retained. IQ has a decreasing effect on probability of being retained in general. Having a young mother increases the probability of being retained at least for the main sample (Table A6 col. (a) and (c)). The variable PR_RET is highly significant and negative for the main sample (Table A6 col. (a), (b), (c)) . However, when we constrain our sample to high school graduates it does not have a significant effect on the probability of being hold in the same grade (Table A7 col. (a), (b), (c)). The variables, DILIG and ABIL, are also most of the time significantly negative. As in Rauber (2007), I also use these variables to measure to what extent a student follows an internal attribution strategy by attributing success to effort and ability. Relying on evidence that individuals with a high degree of self-esteem frequently tend to attribute success as being internal (see Rauber (2007) and its references), the interpretation of the negative coefficients might be that higher self esteem decreases the probability of grade retention. The other variable which is significantly negative for almost all samples is the willingness to pursue higher education (WISH), however with different signs for different subsamples. The coefficient (PARINT) which controls for parents interests on their child's performance at school is for most specifications significantly negative. It means that if parents are more interested in school outcomes, the probability of being retained decreases. For some specifications, the dummy variable for the highest education category of the mother is significant and negative.

In order to evaluate the common support assumption the density of estimated propensity scores by treatment status are drawn for all groups (see Figures from B 1 to B 12). The propensity score graphs do not exhibit a significant common support problem. Nevertheless, I estimate the ATEs twice for each sample. First, I do not apply any common support correction and second I use minima-maxima comparison (see Frölich (2004), Imbens (2004), Imbens and Wooldridge (2007), and Caliendo and Kopeinig (2008)). Minima-maxima comparison is simply discarding the control observations with propensity scores below the minimum propensity score of the treated group and discarding treated observations with propensity scores above the maximum propensity score of the control group.

The estimation results are summarized in Table 1 and 2. Table 1 shows the results for the sample without any restrictions and Table 2 shows the results for the sample of students without previous retention. I estimate the ATE of grade retention for the entire samples and for the subsamples by gender. The estimates of causal effect on high school completion are summarized in the upper panel and the estimates of causal effect on academic grades in the lower panel of Table 1 and 2. The effects are estimated using Doubly Robust Method (DR) (Equation 3.12), weighting by propensity score (PS) (Equation 3.5) and regression (REG) (Equation 3.3) which are outlined in Section 3. For the regression and DR method, the mean functions of the outcome variables are chosen properly according to the features of the outcome variables. The mean function of the binary outcome variable ABI is specified as in Equation 3.9. For the outcome variables MAT and GER, Equation 3.10 is chosen as the mean function. The mean of the last outcome variable GPA is chosen as in Equation 3.8. The control variables are the same as in the propensity score specifications. For each sample, there are two different sets of estimates; column (a) and (b). Column (a) shows the estimation results without applying any common support correction. For the estimates in column (b), I apply minima-maxima comparison to determine the common support. The standard errors are calculated using the asymptotic variance formulas and reported in parentheses.

From Table 1, we see that the effect of grade retention on the probability of completion of upper secondary school for the overall sample is negative according to the DR and REG estimates. The negative effect is higher in magnitude for females than for the entire sample, whereas the effect seems to be positive for males. For all three samples, PS estimates are insignificant. Applying common support restriction only slightly affects the estimates. For the other three outcomes, the estimates by each method are significantly positive for each sample with two exceptions. The PS estimates of the ATE on MAT for females is insignificant with and without common support restriction. The PS estimates of the ATE on GPA for females are insignificant without common support restriction. In the German educational system, grades between 1 and 6 are assigned, where 1 is the best grade and 6 is the worst grade. Therefore, positive estimates of ATE imply a worsening effect on grades. We see that the estimates based on different methods are most of the time very close to each other. The estimates based on DR and REG methods are almost for each case highly significant whereas the PS estimates are sometimes insignificant. It is known that the variance of PS estimates are affected largely by very high and low propensity scores (see for example Khan and Tamer (2007)).

Table 1: Estimated ATE's for the main sample without restrictions according to previous retention.

Outcome	Method	Full Sample		Female		Male	
		(a)	**(b)**	**(a)**	**(b)**	**(a)**	**(b)**
ABI	**DR**	-0.010^{**}	-0.012^{**}	-0.043^{***}	-0.048^{***}	0.017^{***}	0.015^{**}
		(0.006)	(0.005)	(0.013)	(0.012)	(0.007)	(0.007)
	PS	0.002	0.001	-0.026	-0.033	0.018	0.018
		(0.024)	(0.024)	(0.040)	(0.039)	(0.029)	(0.028)
	REG	-0.006	-0.007^{**}	-0.033^{***}	-0.044^{***}	0.014^{***}	0.011^{**}
		(0.004)	(0.004)	(0.006)	(0.007)	(0.005)	(0.005)
number of observations		2726	2711	1257	1200	1469	1436
number of treated		520	519	201	196	319	316
number of untreated		2206	2192	1056	1004	1150	1120
MAT	**DR**	0.266^{***}	0.262^{***}	0.118^{***}	0.123^{***}	0.395^{***}	0.377^{***}
		(0.016)	(0.016)	(0.046)	(0.045)	(0.024)	(0.024)
	PS	0.255^{***}	0.257^{***}	-0.025	0.009	0.405^{***}	0.416^{***}
		(0.067)	(0.066)	(0.105)	(0.104)	(0.083)	(0.083)
	REG	0.271^{***}	0.273^{***}	0.104^{***}	0.109^{***}	0.401^{***}	0.383^{***}
		(0.010)	(0.010)	(0.022)	(0.023)	(0.016)	(0.016)
GER	**DR**	0.296^{***}	0.298^{***}	0.356^{***}	0.358^{***}	0.314^{***}	0.286^{***}
		(0.014)	(0.013)	(0.050)	(0.047)	(0.018)	(0.017)
	PS	0.295^{***}	0.301^{***}	0.225^{**}	0.264^{***}	0.326^{***}	0.341^{***}
		(0.058)	(0.057)	(0.102)	(0.100)	(0.072)	(0.069)
	REG	0.301^{***}	0.300^{***}	0.363^{***}	0.363^{***}	0.308^{***}	0.284^{***}
		(0.008)	(0.008)	(0.022)	(0.022)	(0.013)	(0.013)
GPA	**DR**	0.220^{***}	0.219^{***}	0.177^{***}	0.182^{***}	0.256^{***}	0.242^{***}
		(0.009)	(0.009)	(0.028)	(0.027)	(0.010)	(0.010)
	PS	0.213^{***}	0.219^{***}	0.100	0.135^{*}	0.256^{***}	0.274^{***}
		(0.041)	(0.039)	(0.077)	(0.075)	(0.047)	(0.044)
	REG	0.225^{***}	0.224^{***}	0.189^{***}	0.192^{***}	0.258^{***}	0.245^{***}
		(0.004)	(0.005)	(0.013)	(0.013)	(0.006)	(0.006)
number of observations		1643	1620	686	672	957	922
number of treated		303	299	105	105	198	197
number of untreated		1340	1321	581	567	759	725

The standard errors are calculated as explained in Section 3 and reported in parentheses under the estimates. Column (a) and (b) report the estimates without and with common support restriction respectively. *, **, ***: significant at 10 %, 5 %, 1%

Table 2: Estimated ATE's for the samples without previous retention

Outcome	Method	Full Sample		Female		Male	
		(a)	(b)	(a)	(b)	(a)	(b)
ABI	**DR**	−0.072***	−0.073***	−0.065***	−0.071***	−0.088***	−0.089***
		(0.006)	(0.006)	(0.013)	(0.013)	(0.008)	(0.009)
	PS	−0.062**	−0.059**	−0.049	−0.055	−0.078**	−0.075**
		(0.028)	(0.028)	(0.042)	(0.041)	(0.034)	(0.033)
	REG	−0.075***	−0.075***	−0.064***	−0.068***	−0.090***	−0.092***
		(0.004)	(0.004)	(0.009)	(0.009)	(0.006)	(0.006)
number of observations		1748	1738	866	842	882	850
number of treated		377	377	160	158	217	212
number of untreated		1371	1361	706	684	665	638
MAT	**DR**	0.351***	0.348***	0.143***	0.148***	0.549***	0.542***
		(0.023)	(0.023)	(0.050)	(0.050)	(0.032)	(0.031)
	PS	0.368***	0.363***	0.085	0.108	0.502***	0.541***
		(0.079)	(0.078)	(0.114)	(0.113)	(0.103)	(0.103)
	REG	0.365***	0.367***	0.140***	0.143***	0.546***	0.546***
		(0.012)	(0.012)	(0.027)	(0.028)	(0.018)	(0.019)
GER	**DR**	0.344***	0.342***	0.330***	0.331***	0.402***	0.359***
		(0.024)	(0.024)	(0.052)	(0.051)	(0.026)	(0.024)
	PS	0.385***	0.382***	0.323***	0.348***	0.367***	0.393***
		(0.069)	(0.069)	(0.118)	(0.116)	(0.084)	(0.078)
	REG	0.365***	0.365***	0.352***	0.351***	0.415***	0.371***
		(0.011)	(0.011)	(0.027)	(0.027)	(0.021)	(0.020)
GPA	**DR**	0.243***	0.241***	0.169***	0.170***	0.288***	0.273***
		(0.014)	(0.013)	(0.027)	(0.026)	(0.013)	(0.013)
	PS	0.270***	0.267***	0.199**	0.219***	0.248***	0.284***
		(0.047)	(0.046)	(0.083)	(0.081)	(0.052)	(0.048)
	REG	0.255***	0.256***	0.197***	0.195***	0.292***	0.278***
		(0.006)	(0.006)	(0.016)	(0.016)	(0.009)	(0.009)
number of observations		1248	1242	546	536	702	662
number of treated		227	225	84	84	143	141
number of untreated		1021	1017	462	452	559	521

The standard errors are calculated as explained in Section 3 and reported in parentheses under the estimates. Column (a) and (b) report the estimates without and with common support restriction respectively. *, **, ***: significant at 10 %, 5 %, 1%

Table 2 shows the estimation results for the students who only experienced grade retention after 10^{th} grade. The results are very similar to the previous Table, except that the effect of grade retention on the probability of graduating from high school for male students is significantly negative. Moreover, the estimates are larger in magnitude compared to the previous results. As in previous results, regardless of which method is used the estimates are very close for the same outcome variable. This result should give us some confidence about our model specifications. The negative effect of grade retention on high school completion is higher for boys than girls. Furthermore, the treatment effects on different school grades are also higher for boys than girls. It seems like boys are more negatively affected by grade retention than girls. All in all, our empirical results suggest that grade retention as a school intervention tool does not provide any improvement on average, but has rather worsening effects for students.

5 Conclusion

In this paper, I investigate the causal effect of grade retention on different school outcomes, such as completion of upper secondary school, final grades in math and German as well as the average final grade. The effect of grade retention is an important research topic since at least four decades. The results from previous research are somehow controversial. The literature provides evidence for both negative and positive effects. The methods used for the analysis of the effects range from simple group comparisons to sophisticated econometric modeling. Here, I estimate the effect using a potential outcome framework applying econometric evaluation methods inverse propensity score weighting, regression adjustment and a combination of these two methods. Inverse propensity score weighting estimates are inconsistent if the propensity score is wrongly specified and regression adjustment estimates are inconsistent if the mean function is wrongly specified. Hence, a combination of these two methods gives the researcher some protection against misspecification. The resulting estimator of the ATE is consistent even if only one of the models is correctly specified. An important drawback is that the main underlying assumption, CIA, which provides the identification of the average treatment effect is not testable. As most researchers who uses identification under CIA, I also argue that the rich set of control variables I am using should be enough to satisfy the CIA assumption approximately.

The propensity score estimation results are consistent with much of the existing empirical research on determinants of grade retention. The estimates of the ATE on different outcomes are very close to each other regardless of which of the three methods is chosen. The estimates show that grade retention has a worsening effect on the students' educational achievement. It increases drop-out rate from upper secondary school significantly, and decreases the individual grades in math and German as well as the average final grade. The worsening effect is larger for boys than for girls. Given that grade retention is thought as an intervention tool to improve the educational achievement, our result do not support that this intervention achieves that goal. This result coincides with other empirical results from the US and Canada (see for example Jimerson (1999) and Guevremont, Roos, and Brownell (2007)) and implies the necessity of different approaches to improve the educational achievement.

References

Allensworth, E. (2005): "Dropout Rates After High-Stakes Testing in Elementary School: A Study of the Contradictory Effects of Chicagos Efforts to End Social Promotion," *Educational Evaluation and Policy Analysis*, 27, 341364.

Amthauer, R. (1953): *Intelligenz-Struktur-Test*. Verlag fr Psychologie, Dr. C.J. Hogrefe, Gttingen, 2. erweiterte auflage edn.

Balitewicz, T. F. (1998): "The Long-Term Effects of Grade Retention," Discussion paper, ERIC Document No. 424 616.

Bang, H., and J. M. Robins (2005): "Doubly Robust Estimation in Missing Data and Causal Inference Models," *Biometrics*, 61, 962–972.

Beebe-Frankenberger, M., K. M. Bocian, D. L. MacMillan, and F. M. Gresham (2004): "Sorting Second-Grade Students: Differentiating Those Retained from Those Promoted," *Journal of Educational Psychology*, 96, 204215.

Blair, C. (2001): "The Early Identification of Risk for Grade Retention Among African American Children at Risk for School Difficulty," *Applied Developmental Science*, 5, 3750.

Byrd, R. S., and M. L. Weitzman (1994): "Predictors of Early Grade Retention Among Children in the United States," *Pediatrics*, 93, 481487.

CAESR (2007): *Dataset Gymnasiastenstudie*. Central Archive for Empirical Social Research, Köln.

Caliendo, M., and S. Kopeinig (2008): "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of Economic Surveys*, 22, 31–72.

Corman, H. (2003): "The Effects of State Policies, Individual Characteristics, Family Characteristics, and Neighbourhood Characteristics on Grade Repetition in the United States," *Economics of Education Review*, 22, 409420.

Dauber, S. L., K. L. Alexander, and D. R. Entwisle (1993): "Characteristics of Retainees and Early Precursors of Retention in Grade: Who Is Held Back?," *Merrill-Palmer Quarterly*, 39, 326343.

Eide, E. R., and M. H. Showalter (2001): "The Effect of Grade Retention on Education and Labor Market Outcomes," *Economics of Education Review*, 20, 563–576.

El-Hassan, K. (1998): "Relation of Academic History and Demographic Variables to Grade Retention in Lebanon," *Journal of Educational Research*, 91, 279288.

Ferguson, P., S. R. Jimerson, and M. J. Dalton (2001): "Sorting Out Successful Failures: Exploratory Analyses of Factors Associated with Academic and Behavioral Outcomes of Retained Students," *Psychology in the Schools*, 38, 327341.

Fine, J. G., and J. M. Davis (2003): "Grade Retention and Enrollment in Post-Secondary Education," *Journal of School Psychology*, 41, 401411.

FREDERICK, C. B., AND R. M. HAUSER (2008): "Have We Put an End to Social Promotion? Changes in School Progress Among Children Aged 6 to 17 from 1972 to 2005," *Demography*, 45, 719740.

FRÖLICH, M. (2004): "A Note on the Role of the Propensity Score for Estimating Average Treatment Effects," *Econometric Reviews*, 23(2), 167 – 174.

FRYMIER, J. (1997): "Characteristics of Students Retained in Grade," *The High School Journal*, 80, 184192.

GOLDSCHMIDT, P., AND J. WANG (1999): "When Can Schools Affect Dropout Behavior? A Longitudinal Multilevel Analysis," *American Educational Research Journal*, 36, 715738.

GREENE, J. P., AND M. A. WINTERS (2004): "An Evaluation of Floridas Program to End Social Promotion," Education Working Paper No. 7, Center for Civic Innovation, Manhattan Institute for Policy Research.

——— (2007): "Revisiting Grade Retention: An Evaluation of Floridas Test- Based Promotion Policy," *Education Finance and Policy*, 2, 319340.

——— (2009): "The Effects of Exemptions to Floridas Test-Based Promotion Policy: Who Is Retained? Who Benefits Academically?," *Economics of Education Review*, 28, 135142.

GUEVREMONT, A., N. P. ROOS, AND M. BROWNELL (2007): "Predictors and Consequences of Grade Retention: Examining Data from Manitoba, Canada," *Canadian Journal of School Psychology*, 22, 50–67.

HIRANO, K., AND G. IMBENS (2001): "Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization," *Health Services and Outcomes Research Methodology*, 2, 259–278.

HIRANO, K., G. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71.

HONG, G., AND B. YU (2007): "Early-Grade Retention and Childrens Reading and Math Learning in Elementary Years," *Educational Evaluation and Policy Analysis*, 29, 239261.

HORVITZ, D. G., AND D. J. THOMPSON (1952): "A Generalization of Sampling Without Replacement From a Finite Universe," *Journal of the American Statistical Association*, 47, 663–685.

IMBENS, G. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity," *Review of Economics and Statistics*, 86, 4–29.

IMBENS, G., AND J. WOOLDRIDGE (2007): "What is New in Econometrics, Lecture 1, Estimation of Average Treatment Effects under Unconfoundedness," NBER Lectures.

JACKSON, G. B. (1975): "The Research Evidence on the Effects of Grade Retention," *Review of Educational Research*, 45, 613–635.

Jacob, B. A., and L. Lefgren (2002): "Remedial Education and Student Achievement: A Regression- Discontinuity Analysis," NBER Working Paper 8918, Cambridge, MA: National Bureau of Economic Research.

——— (2007): "The Effect of Grade Retention on High School Completion," NBER Working Paper 13514, Cambridge, MA: National Bureau of Economic Research.

Jimerson, S., E. Carlson, M. Rotert, B. Egeland, and L. A. Sroufe (1997): "A Prospective, Longitudinal Study of the Correlates and Consequences of Early Grade Retention," *Journal of School Psychology*, 35, 3–25.

Jimerson, S. R. (1999): "On the Failure of Failure: Examining the Association Between Early Grade Retention and Education and Employment Outcomes During Late Adolescence," *Journal of School Psychology*, 37, 243272.

——— (2001): "Meta-Analysis of Grade Retention Research: Implications for Practice in the 21st Century," *School Psychology Review*, 30, 420438.

Jimerson, S. R., G. E. Anderson, and A. D. Whipple. (2002): "Winning the Battle and Losing the War: Examining the Relation Between Grade Retention and Dropping Out of High School," *Psychology in the Schools*, 39, 441457.

Jimerson, S. R., P. Ferguson, A. D. Whipple, G. E. Anderson, and M. J. Dalton (2002): "Exploring the Association Between Grade Retention and Dropout: A Longitudinal Study Examining Socio- Emotional, Behavioral, and Achievement Characteristics of Retained Students," *The California School Psychologist*, 7, 5162.

Khan, S., and E. Tamer (2007): "Irregular Identification, Support Conditions, and Inverse Weight Estimation," Discussion paper, Unpublished manuscript, Northwestern University 2009.

Liddell, C., and G. Rae (2001): "Predicting Early Grade Retention: A Longitudinal Investigation of Primary School Progress in a Sample of Rural South African Children," *Psychology*, 71, 413428.

McArthur, E. K., and S. M. Bianchi (1993): "Characteristics of children who are behind in school.," National Center for Education Statistics and Bureau of the Census, ERIC Document Reproduction.

McCoy, A. R., and A. J. Reynolds (1999a): "Grade Retention and School Performance: An Extended Investigation," *Journal of School Psychology*, 37, 273–298.

——— (1999b): "Grade Retention and School Performance: An Extended Investigation," *Journal of School Psychology*, 37, 273298.

Rauber, M. (2007): "Noncognitive Skills and Success in Life: The Importance of Motivation and Self- Regulation," Discussion Paper 07.

Robins, J. M., and Y. Ritov (1997): "Toward a Curse of Dimensionality Appropriate (CODA) Asymptotic Theory for Semi-Parametric Models," *Statistics in Medicine*, 16, 285–319.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1995): "Analysis of Semiparametric Regression Models for Repeated Outcomes under the Presence of Missing Data," *Journal of the American Statistical Association*, 90, 106–121.

Robles-Pina, R. A., E. Defrance, and D. L. Cox (2008): "Self-Concept, Early Childhood Depression and School Retention as Predictors of Adolescent Depression in Urban Hispanic Adolescents," *School Psychology International*, 29, 426441.

Roderick, M. (1994): "Grade Retention and School Dropout: Investigating the Association," *American Educational Research Journal*, 31, 729759.

Rosenbaum, P. R. (1987): "Model-Based Direct Adjustment," *Journal of the American Statistical Association*, 82, 387–394.

Rosenbaum, P. R., and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41 – 55.

Scharfstein, D. O., A. Rotnitzky, and J. M. Robins (1999): "Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion)," *Journal of the American Statistical Association*, 94, 1096–1120.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

——— (2007): "Inverse Probability Weighted Estimation for General Missing Data Problems," *Journal of Econometrics*, 141, 1281 – 1301.

——— (2009): "Average Treatment Effect Estimation: Unconfounded Treatment Assignment," Unpublished Manuscript, Michigan State University.

Xia, N., and S. N. Kirby (2009): "Retaining Students in Grade: A Literature Review of the Effects of Retention on Students' Academic and Nonacademic Outcomes," RAND Cooperation Technical Report.

A Tables

Table A1: Definition of Variables

Variable	Definition
ABI	Dummy, 1 if upper secondary school degree held (Abitur)
MAT	Grade in math in the last year of upper secondary school between 1-6, 1 is the best grade
GER	Grade in German in the last year of upper secondary school between 1-6, 1 is the best grade
RET	Dummy, 1 if a grade is repeated at least once in the school year 1970/71 or later
SHNR	School number
FEMALE	Dummy, 1 if female
AGE	Age in years
IQ	Number of correctly solved questions in the Intelligence Structure Test (IST; Amthauer (1953)). The test was carried out in 1969.
EDU_MOT	Categorical variable for educational attainment of the mother from 1-4
EDU_MOTj	Dummy, 1 if EDU_MOT=j for $j = 1, 2, 3, 4$
EDU_FAT	Categorical variable for educational attainment of the father from 1-4
EDU_FATj	Dummy, 1 if EDU_FAT=j for $j = 1, 2, 3, 4$
HHINC	Categorical variable for net household income in 1970 from 1-9 =1 up to 750 DM, =2 751 up to 1000 DM, =3 1001 up to 1250 DM, =4 1251 up to 1500 DM, =5 1501 up to 2000 DM,=6 2001 up to 2500 DM, =7 2501 up to 3000 DM, =8 3001 up to 4000 DM,=9 more than 4000 DM
EMP_MOT	Categorical variable for mother's employment status from 1-3
EMP_MOT1	Dummy, 1 if the mother is employed during the survey (EMP_MOT=1)
EMP_MOT2	Dummy, 1 if the mother is unemployed, but was employed before the survey (EMP_MOT=2)
EMP_MOT3	Dummy, 1 if the mother is out of labour force (EMP_MOT=3)
PARINT1	Dummy, 1 if parents are interested in promotion on to the next grade level
PARINT2	Dummy, 1 if parents are interested in final grades
PARINT3	Dummy, 1 if parents are interested in test grades
INTSCHOOL	Average value of PARINT1, PARINT2 and PARINT3
AGEMOT	Categorical variable for mother's age from 1-9 =1 if 30-34, =2 if 35-39, =3 if 40-44, =4 if 45-49, =5 if 50-54, =6 if 55-59, =7 if 60-64, =8 if 65-70, =9 if she died
AGEMOTj	Dummy, 1 if AGEMOT=j
WISH	Do you want to continue studying after upper secondary school? =1 if the answer is yes, =2 if maybe, =3 if no, =4 if do not know yet, =5 if no upper secondary school degree is planned
WISHj	Dummy, =1 if WISH=j
PR_RET	Dummy, 1 if a grade is repeated at least once before the school year 1969/70
DILIG	Measure of attributing success to diligence on a scale from 0 (weaker) to 5 (stronger)
ABIL	Measure of attributing success to ability on a scale from 0 (weaker) to 5 (stronger)

Source: Dataset Gymnasiastenstudie, own definitions

Table A2: Summary Statistics of Unrestricted Sample (Sample 1)

Variable	Mean	Std Dev	Minimum	Maximum
ABI	0.64	0.48	0	1
RET	0.19	0.39	0	1
FEMALE	0.46	0.50	0	1
AGE	15.41	0.90	13	19
IQ	40.72	8.94	12	70
EDU_MOT	4.19	3.50	1	13
EDU_VAT	5.86	4.26	1	13
HHINC	4.50	2.06	1	9
EMP_MOT	2.02	0.70	1	3
PARINT1	0.64	0.48	0	1
PARINT2	0.61	0.49	0	1
PARINT3	0.75	0.43	0	1
INTSCHOOL	0.67	0.30	0	1
AGE_MOT	3.61	1.18	1	9
PR_RET	0.36	0.48	0	1
WISH	2.62	1.56	1	5
DILIG	4.12	1.04	0	5
ABIL	3.51	1.08	0	5
Number of Observations			2726	

Sample: Sample without restrictions on previous grade retention

Table A3: Summary Statistics of Sample 2

Variable	Mean	Std	Minimum	Maximum
GPA	2.97	0.54	1.08	4.10
MAT	3.48	1.08	1	6
GER	3.33	0.85	1	5
RET	0.18	0.39	0	1
FEMALE	0.42	0.49	0	1
AGE	15.19	0.82	13	19
IQ	41.78	9.12	15	70
EDU_MOT	4.36	3.60	1	13
EDU_VAT	6.10	4.31	1	13
HHINC	4.57	2.06	1	9
EMP_MOT	2.03	0.69	1	3
PARINT1	0.64	0.48	0	1
PARINT2	0.64	0.48	0	1
PARINT3	0.78	0.41	0	1
INTSCHOOL	0.69	0.30	0	1
AGE_MOT	3.62	1.18	1	9
PR_RET	0.24	0.43	0	1
WISH	2.15	1.32	1	5
DILIG	4.09	1.05	0	5
ABIL	3.51	1.09	0	5
Number of Observations			1643	

Sample: Graduates from upper secondary school. Sample 1 restricted by ABI=1

Table A4: Summary Statistics of Sample 3

Variable	Mean	Std. Dev.	Minimum	Maximum
ABI	0.75	0.43	0	1
RET	0.22	0.41	0	1
FEMALE	0.50	0.50	0	1
AGE	15.03	0.71	13	19
IQ	40.85	9.08	13	70
EDU_MOT	4.15	3.55	1	13
EDU_VAT	5.76	4.27	1	13
HHINC	4.41	2.07	1	9
EMP_MOT	2.03	0.69	1	3
PARINT1	0.64	0.48	0	1
PARINT2	0.63	0.48	0	1
PARINT3	0.77	0.42	0	1
INTSCHOOL	0.68	0.30	0	1
AGE_MOT	3.53	1.16	1	9
WISH	2.45	1.50	1	5
DILIG	4.11	1.03	0	5
ABIL	3.57	1.05	0	5
Number of Observations			1748	

Sample: Students without previous grade retention. Sample 1 restricted by PR_RET=0

Table A5: Summary Statistics of Sample 4

Variable	Mean	Std. Dev.	Minimum	Maximum
GPA	2.92	0.54	1.08	4.00
MAT	3.36	1.09	1	6
GER	3.26	0.87	1	5
RET	0.18	0.39	0	1
FEMALE	0.44	0.50	0	1
AGE	14.96	0.69	13	19
IQ	41.79	9.22	15	70
EDU_MOT	4.26	3.62	1	13
EDU_VAT	5.92	4.29	1	13
HHINC	4.46	2.06	1	9
EMP_MOT	2.03	0.68	1	3
PARINT1	0.65	0.48	0	1
PARINT2	0.79	0.41	0	1
PARINT3	2.12	1.30	1	5
INTSCHOOL	0.69	0.30	0	1
AGE_MOT	3.56	1.17	1	9
WISH	2.12	1.30	1	5
DILIG	4.10	1.04	0	5
ABIL	3.56	1.06	0	5
Number of Observations			1248	

Sample: Graduates from upper secondary school without previous grade retention. Sample 3 restricted by ABI=1

Table A6: Propensity Score Estimation Results for Sample 1 and Sample 2

	DATA 1			DATA 2		
	(a)	(b)	(c)	(d)	(e)	(f)
Variable	Full sample	Female	Male	Full sample	Female	Male
Constant	3.060***	3.067	2.506*	3.750**	2.870	3.575*
	1.242	2.091	1.555	1.680	3.101	2.073
SHNR	−0.006**	−0.009**	-0.004	−0.009***	−0.015***	-0.006
	0.003	0.004	0.003	0.003	0.006	0.004
FEMALE	−0.534***	–	–	−0.604***	–	–
	0.110			0.146		
AGE	-0.093	-0.103	-0.070	0.013	0.079	0.013
	0.070	0.122	0.087	0.093	0.179	0.111
IQ	−0.034***	−0.039***	−0.031***	−0.039***	−0.063***	−0.027***
	0.006	0.010	0.008	0.008	0.014	0.010
EDU_MOT2	-0.060	-0.039	-0.046	-0.127	-0.020	-0.206
	0.141	0.218	0.188	0.188	0.312	0.241
EDU_MOT3	-0.105	-0.136	-0.140	-0.272	-0.455	-0.309
	0.207	0.325	0.274	0.278	0.482	0.354
EDU_MOT4	-0.308	0.262	−0.777**	-0.453	0.370	-1.045
	0.243	0.358	0.337	0.319	0.474	0.445
EDU_VAT2	0.257*	0.385*	0.122	0.161	0.091	0.158
	0.155	0.230	0.216	0.214	0.343	0.279
EDU_VAT3	0.102	0.202	0.052	0.203	0.308	0.170
	0.164	0.258	0.216	0.210	0.361	0.264
EDU_VAT4	0.217	0.074	0.348	0.324	0.261	0.472
	0.190	0.302	0.249	0.249	0.403	0.322
HHINC	-0.036	−0.088*	0.000	-0.064	−0.132*	-0.028
	0.031	0.050	0.041	0.042	0.072	0.053
EMP_MOT1	0.315**	0.254	0.326*	0.305	0.141	0.318
	0.144	0.244	0.183	0.190	0.344	0.235
EMP_MOT2	0.069	0.158	0.035	-0.036	-0.020	-0.017
	0.125	0.204	0.161	0.161	0.274	0.204
INTSCHOOL	−0.384**	−0.852***	-0.070	-0.247	-0.586	-0.066
	0.169	0.277	0.219	0.219	0.379	0.275
AGEMOT2	-0.451	-0.557	-0.386	-0.590	-0.325	-0.632
	0.420	0.714	0.527	0.538	1.147	0.638
AGEMOT3	−0.671*	-0.538	-0.805	-0.570	-0.115	-0.734
	0.410	0.700	0.514	0.519	1.118	0.614
AGEMOT4	−0.735*	-0.696	-0.825	-0.733	-0.155	-0.993
	0.412	0.706	0.516	0.522	1.122	0.619
AGEMOT5	−0.741*	-0.650	-0.833	-0.874	-0.433	-1.019
	0.432	0.733	0.543	0.550	1.170	0.650
AGEMOT6	−1.225***	-0.840	−1.509**	−1.132*	-0.401	−1.385*
	0.492	0.805	0.635	0.612	1.232	0.741
AGEMOT7	−1.458*	-0.801	-1.794	-1.540	(omitted)	-1.483
	0.865	1.284	1.188	1.228		1.272
AGEMOT8	−1.612*	(omitted)	-1.424	-2.035	(omitted)	-1.775
	0.871		0.959	1.185		1.256
PR_RET	−0.414***	−0.439*	−0.414***	-0.037	-0.200	-0.002
	0.133	0.230	0.167	0.179	0.329	0.219
WISH1	0.422**	0.684***	0.176	−1.509***	−1.276**	−1.876***
	0.190	0.267	0.277	0.391	0.538	0.636
WISH2	0.537***	0.932***	0.191	−1.308***	-0.844	−1.768***
	0.204	0.285	0.301	0.402	0.553	0.653
WISH3	0.644***	0.482	0.664*	−1.272***	−1.368*	−1.336*
	0.256	0.426	0.347	0.465	0.737	0.706
WISH4	0.788***	0.829***	0.691***	−1.105***	−1.045**	−1.383**
	0.187	0.265	0.275	0.390	0.538	0.637
DILIG	−0.076*	-0.086	-0.071	-0.129	-0.098	−0.145**
	0.047	0.081	0.059	0.061	0.115	0.073
ABIL	−0.104**	-0.059	−0.138**	−0.111*	-0.085	−0.153**
	0.046	0.077	0.058	0.059	0.103	0.074
No. of Obs.	2726	1249	1469	1643	680	957
Log-likelihood	-1258.31	-515.10	-729.95	-740.46	-267.99	-461.91
LR chi2(k)	140.22	71.92	77.50	89.89	49.20	51.97

The standard errors are reported in parentheses under the estimates. *, **, ***: significant at 10 %, 5 %, 1%

Table A7: Propensity Score Estimation Results for Different Samples data2

	DATA 3			DATA 4		
	(a)	(b)	(c)	(d)	(e)	(f)
Variable	Full sample	Female	Male	Full sample	Female	Male
Constant	1.049**	-1.096	2.550	2.290	-1.913	4.209
	1.545	2.383	2.117	2.061	3.578	2.730
SHNR	−0.006**	-0.006	-0.006	−0.009**	-0.012	-0.008
	0.003	0.005	0.004	0.004	0.006	0.005
FEMALE	−0.561***	–	–	−0.624***	–	–
	0.131			0.168		
AGE	0.062	0.154	-0.004	0.111	0.362*	0.025
	0.087	0.140	0.116	0.113	0.208	0.139
IQ	−0.034***	−0.030***	−0.037***	−0.035***	−0.055***	−0.027**
	0.007	0.011	0.010	0.009	0.015	0.011
EDU_MOT2	-0.026	0.010	-0.051	-0.118	-0.028	-0.186
	0.172	0.251	0.243	0.223	0.355	0.295
EDU_MOT3	-0.061	-0.099	-0.071	-0.255	-0.488	-0.156
	0.256	0.368	0.370	0.329	0.529	0.446
EDU_MOT4	-0.435	-0.033	−0.817*	-0.634	0.123	−1.151**
	0.302	0.431	0.431	0.391	0.566	0.568
EDU_VAT2	0.395**	0.492*	0.271	0.360	0.263	0.375
	0.185	0.260	0.274	0.246	0.388	0.327
EDU_VAT3	0.088	0.230	0.040	0.359	0.465	0.352
	0.200	0.297	0.278	0.246	0.409	0.317
EDU_VAT4	0.272	0.171	0.356	0.444	0.518	0.443
	0.238	0.351	0.337	0.305	0.458	0.424
HHINC	-0.013	-0.069	0.040	-0.017	-0.077	0.024
	0.037	0.057	0.051	0.048	0.083	0.062
EMP_MOT1	0.544***	0.487*	0.570	0.479**	0.450	0.466*
	0.175	0.280	0.232	0.221	0.396	0.276
EMP_MOT2	0.173	0.259	0.157	0.040	0.241	-0.038
	0.153	0.236	0.205	0.191	0.320	0.244
INTSCHOOL	−0.525***	−0.904***	-0.277	-0.391	−0.761*	-0.183
	0.201	0.314	0.273	0.253	0.427	0.326
AGEMOT2	-0.291	-0.834	0.160	-0.293	-0.517	-0.271
	0.534	0.739	0.786	0.715	1.163	0.934
AGEMOT3	-0.525	-0.791	-0.346	-0.317	-0.458	-0.373
	0.526	0.724	0.775	0.701	1.137	0.918
AGEMOT4	-0.658	-1.096	-0.368	-0.476	-0.600	-0.559
	0.529	0.735	0.778	0.704	1.146	0.922
AGEMOT5	-0.555	-1.001	-0.215	-0.382	-0.644	-0.370
	0.552	0.769	0.808	0.733	1.201	0.952
AGEMOT6	-0.661	-0.914	-0.469	-0.441	-0.637	-0.241
	0.605	0.831	0.892	0.784	1.258	1.032
AGEMOT7	-1.702	(omitted)	-0.989	-1.194	(omitted)	-0.967
	1.208		1.374	1.344		1.462
AGEMOT8	-1.028	(omitted)	-0.688	-1.197	(omitted)	-1.070
	0.973		1.168	1.304		1.460
WISH1	-0.166	0.279	−0.874**	−2.618***	−2.125***	−3.517***
	0.232	0.301	0.399	0.527	0.670	1.120
WISH2	0.182	0.625**	-0.537	−2.271***	−1.790***	−3.117***
	0.242	0.317	0.415	0.535	0.683	1.128
WISH3	0.342	0.193	0.018	−2.206***	−2.161***	−2.760**
	0.313	0.516	0.474	0.598	0.879	1.171
WISH4	0.309	0.562*	-0.230	−2.065***	−1.760***	−2.866***
	0.226	0.293	0.396	0.523	0.665	1.118
DILIG	−0.104*	-0.084	-0.118	-0.073	-0.013	-0.096
	0.057	0.091	0.075	0.073	0.135	0.089
ABIL	−0.123**	0.004	−0.219***	-0.088	0.044	−0.188**
	0.056	0.088	0.076	0.071	0.117	0.091
No. of Obs.	1748	860	882	1248	543	702
Log-likelihood	-863.10	-393.98	-457.55	-551.42	-214.12	-329.11
LR chi2(k)	96.54	38.40	69.10	80.88	39.57	51.50

The standard errors are reported in parentheses under the estimates. *, **, ***: significant at 10 %, 5 %, 1%

B Figures

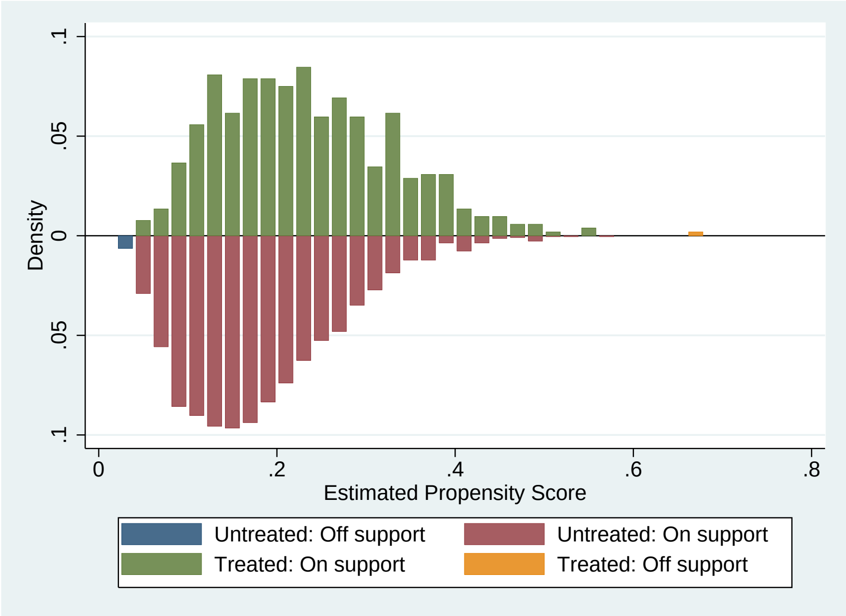

Figure B1: Density of Estimated Probability of Grade Retention for Sample 1. Estimation is based on specification given in Table A6, Col. (a)

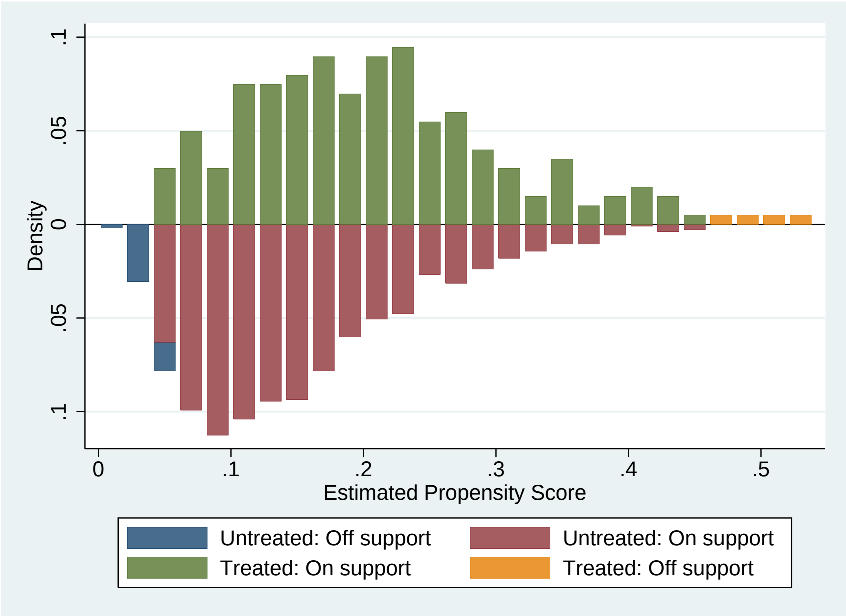

Figure B2: Density of Estimated Probability of Grade Retention for females of Sample 1. Estimation is based on specification given in Table A6, Col. (b)

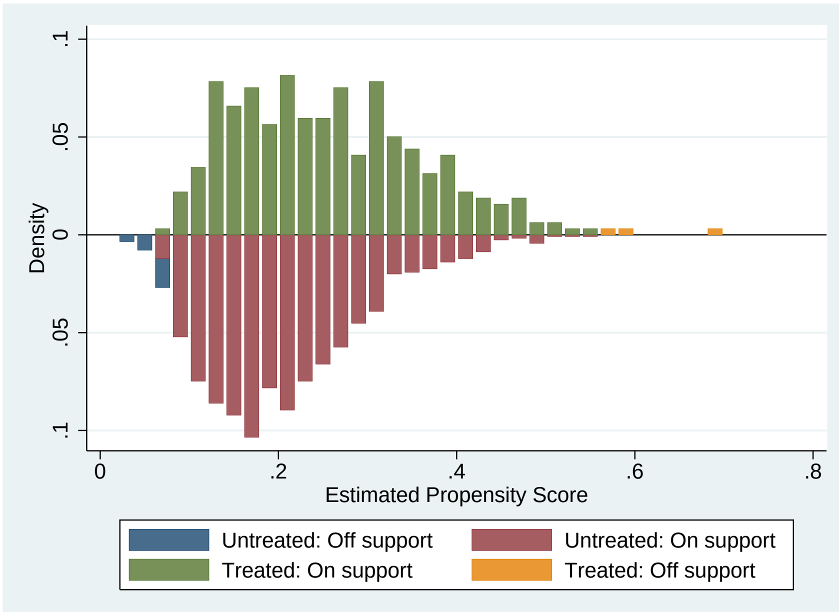

Figure B3: Density of Estimated Probability of Grade Retention for males of Sample 1. Estimation is based on specification given in Table A6, Col. (c)

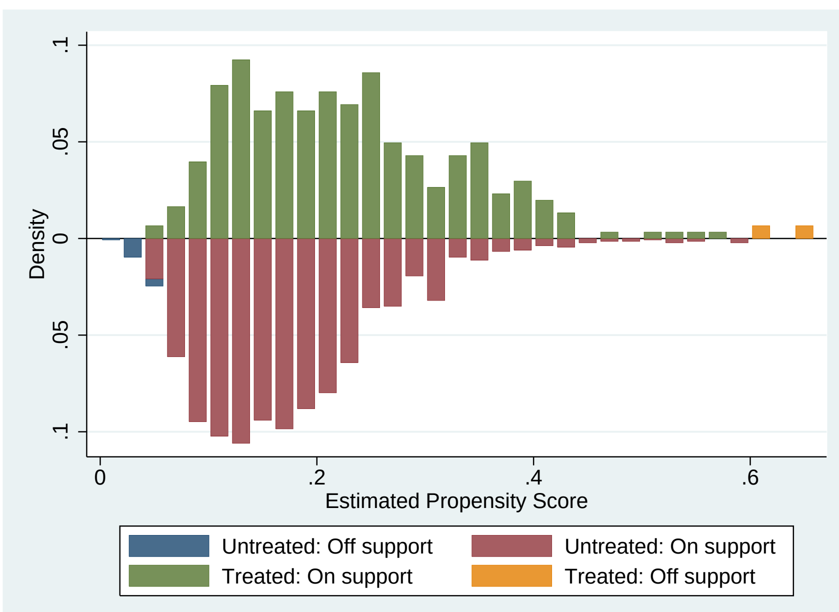

Figure B4: Density of Estimated Probability of Grade Retention for Sample 2. Estimation is based on specification given in Table A6, Col. (d)

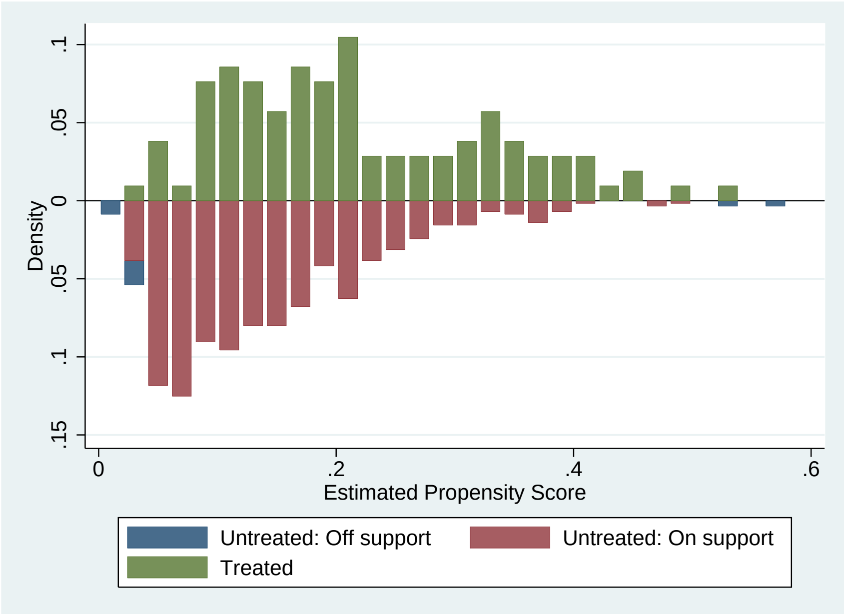

Figure B5: Density of Estimated Probability of Grade Retention for females of Sample 2. Estimation is based on specification given in Table A6, Col. (e)

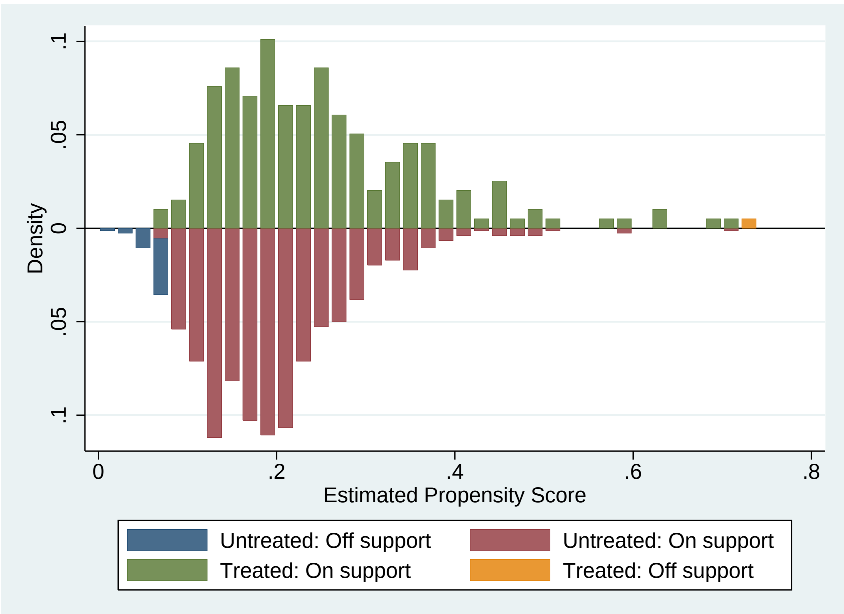

Figure B6: Density of Estimated Probability of Grade Retention for males of Sample 2. Estimation is based on specification given in Table A6, Col. (f)

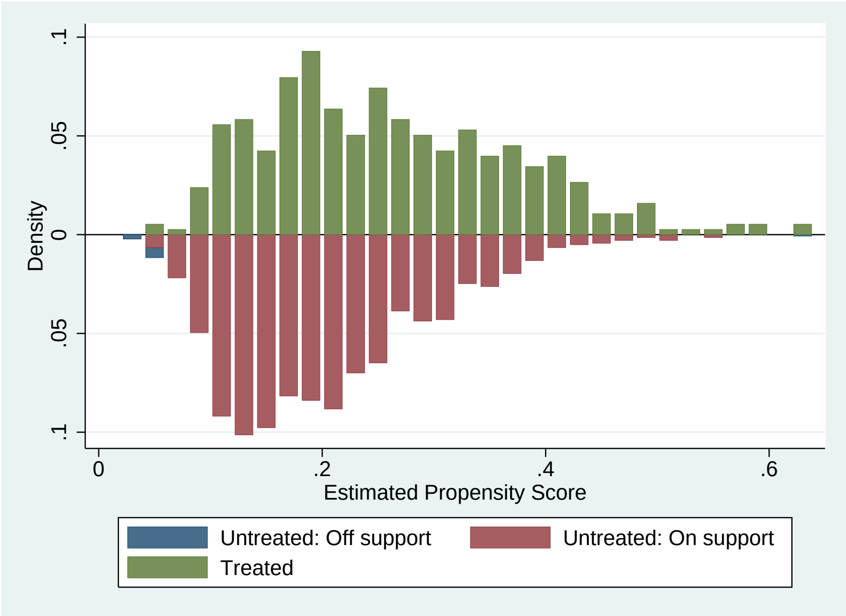

Figure B7: Density of Estimated Probability of Grade Retention for Sample 3. Estimation is based on specification given in Table A7, Col. (a)

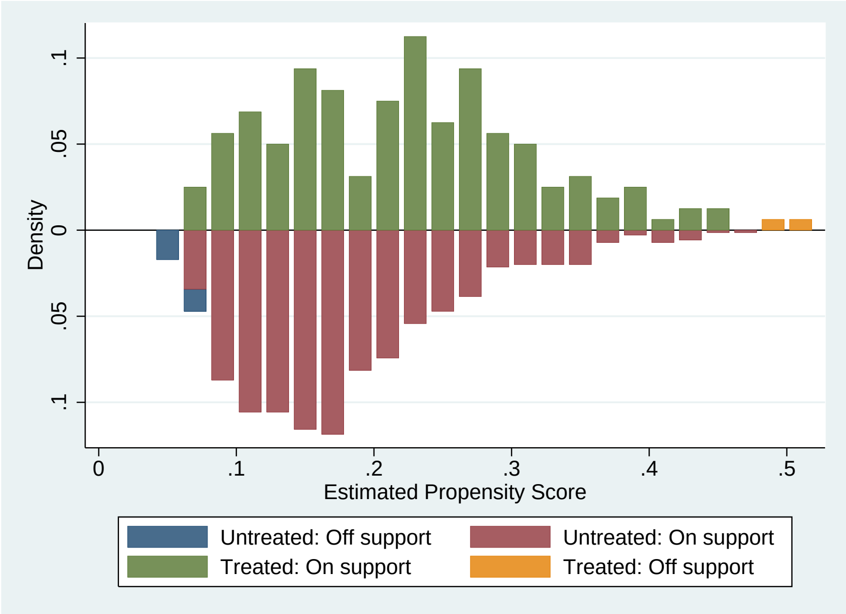

Figure B8: Density of Estimated Probability of Grade Retention for females of Sample 3. Estimation is based on specification given in Table A7, Col. (b)

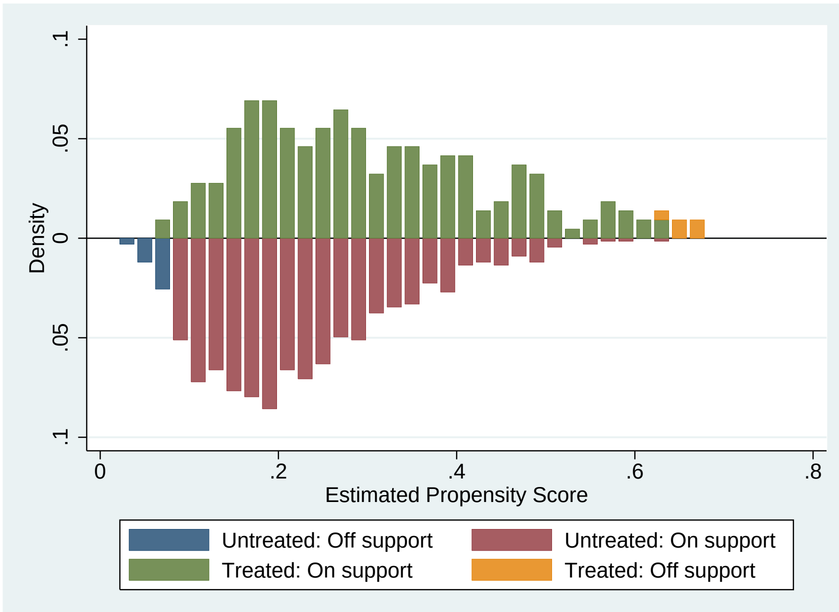

Figure B9: Density of Estimated Probability of Grade Retention for females of Sample 3. Estimation is based on specification given in Table A7, Col. (c)

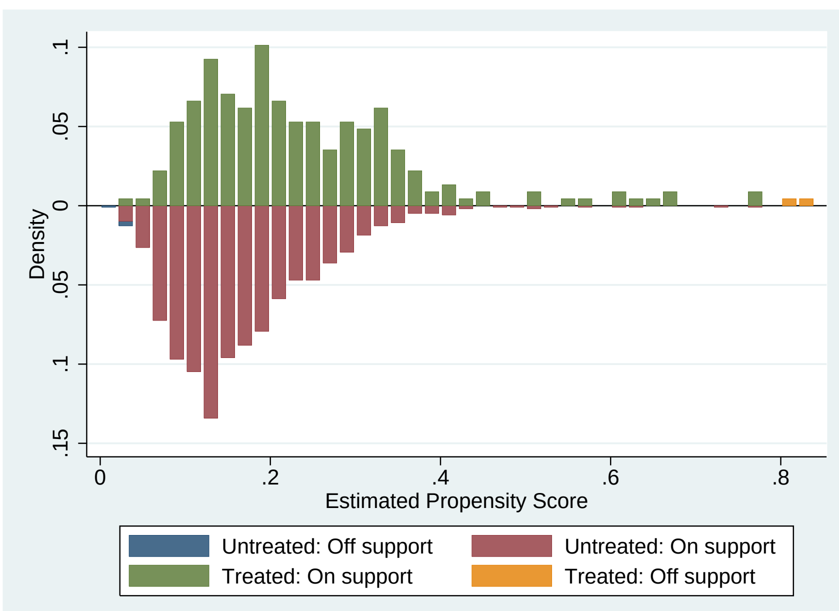

Figure B10: Density of Estimated Probability of Grade Retention for Sample 4. Estimation is based on specification given in Table A7, Col. (d)

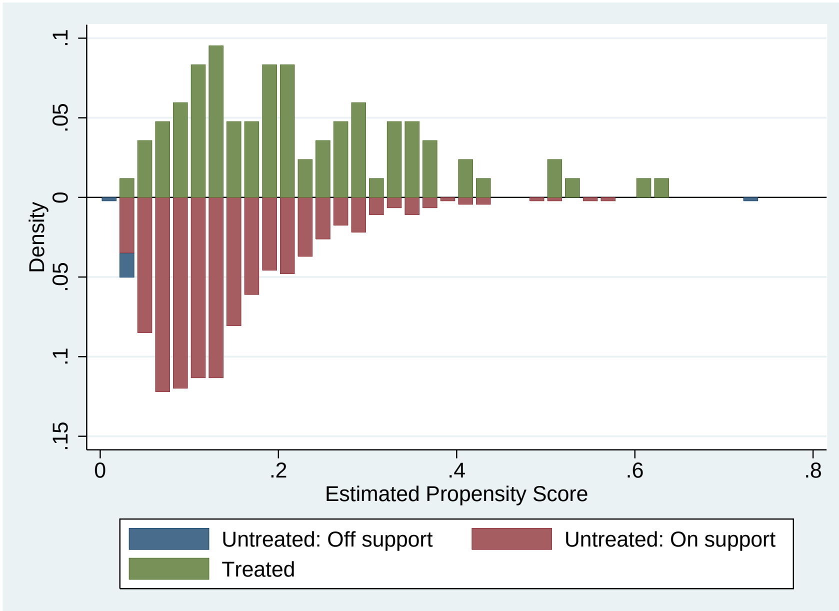

Figure B11: Density of Estimated Probability of Grade Retention for females of Sample 4. Estimation is based on specification given in Table A7, Col. (e)

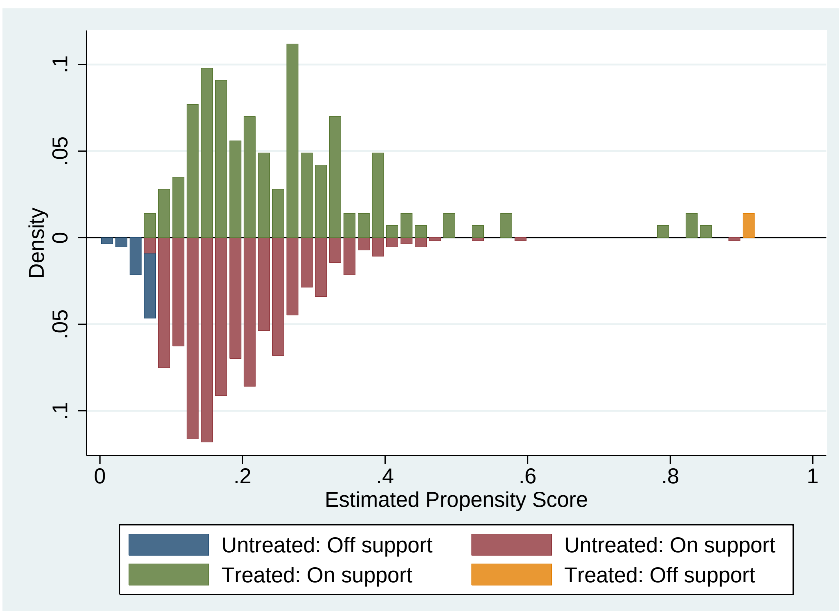

Figure B12: Density of Estimated Probability of Grade Retention for males of Sample 4. Estimation is based on specification given in Table A7, Col. (f)

CMS Discussion Paper 2010-01

For a complete list of the Discussion Papers published by the CMS please visit http://www.uni-konstanz.de/cms.

01 "The Effect of Grade Retention on School Outcomes: An Application of Doubly Robust Estimation Method" by Selver Derya Uysal, January 2010